Assembling the human genome using long nanopore sequencing reads

To gain a comprehensive insight into human genetic variation, and its potential impact on disease risk, it is important to obtain fully characterised, complete genomes. However, the presence of large structural variants (SVs) and repeat sequences have posed a significant challenge to assembling the human genome to completion.

Unlike short-read sequencing, nanopore sequencing produces long and ultra-long reads, which enhance the resolution of SVs and repeats. Epigenetic modifications can also be explored through direct sequencing of native DNA. With high-yield PromethION™ devices, sequencing and assembling highly contiguous human genomes is now possible, with unprecedented efficiency.

Here we present a simple workflow for human genome assembly from a blood sample, using the PromethION sequencing device range.

EXTRACTION: obtaining high molecular-weight DNA

View more extraction protocol recommendations for your sample type, plus guidance on DNA storage and contaminants: **community.nanoporetech.com/docs/prepare**

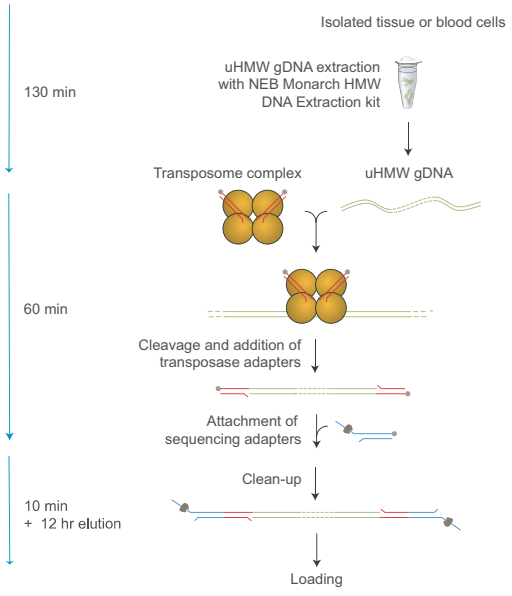
Selecting a suitable extraction method for obtaining high molecular-weight DNA greatly depends on sample type. For the extraction of ultra-high molecular-weight DNA from whole blood, we recommend using the **NEB Monarch HMW DNA Extraction Kit**.

However, if it is not possible to use this option — for example, if the starting sample is too fragmented or insufficient sample is available — we recommend using the **QIAGEN Puregene Blood Kit**, which we have found maximises the production of 25–35kb sequencing reads.

LIBRARY PREPARATION: selecting a kit

Find out more about library prep solutions: **nanoporetech.com/products/kits**

Enriching for long and ultra-long (≥50 kb) gDNA fragments is important for performing genome assembly, to maximise the overlap of sequencing reads in analysis. Generating ultra-long reads will ensure the greatest contiguity and continuity for assembly. We recommend preparing ultra-high molecular-weight gDNA for sequencing using the **Ultra-Long DNA Sequencing Kit** — generating read length N50s >50 kb. However, if it is not possible to isolate high molecular-weight DNA in extraction, we recommend using the **Ligation Sequencing Kit**, which provides the greatest yield and control over read lengths. When using this kit, we recommend size selection and light shearing of the extracted gDNA, which we have found to improve the read length N50. The Oxford Nanopore **Short Fragment Eliminator Expansion Kit** is recommended to size select for fragments >25 kb, and the **Diagenode Megaruptor 3** is recommended for shearing.

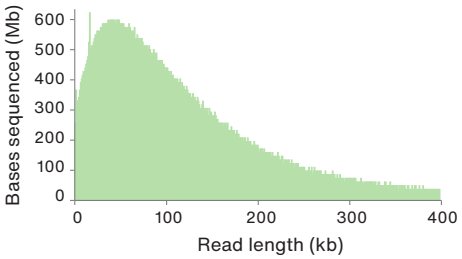


SEQUENCING: generating high yields of long reads with PromethION

Find out more about PromethION:
nanoporetech.com/products/promethion

To assemble a human genome, we recommend sequencing to a minimum depth of 20x when using the **Ultra-Long DNA Sequencing Kit**. This can be achieved by sequencing a single genomic sample on one PromethION Flow Cell. For best assembly metrics, sequencing to a depth of 30x will further improve completeness and contiguity. Throughput can be maximised by washing the flow cell using the **Flow Cell Wash Kit** and loading fresh library every 24 hours. We recommend basecalling using high accuracy mode or super accuracy mode.

Read length distribution obtained from a library (mammalian blood) using the Ultra-Long DNA Sequencing Kit



If using the **Ligation Sequencing Kit**, we recommend sequencing to a minimum depth of 30x of 25–35 kb reads. This can also be performed on one PromethION Flow Cell.

The high-throughput PromethION 24 and 48 sequencing devices have the capacity to run up to 24 or 48 high-yield PromethION Flow Cells, providing ultimate flexibility and adaptability to your sequencing needs. For lower throughput requirements, the compact PromethION 2 Solo – which can be plugged into a GridION™, and the standalone PromethION 2 enable sequencing on up to two flow cells, for PromethION-scale sequencing in any lab.

Find out more about the Flow Cell Wash Kit: **store.nanoporetech.com/flow-cell-wash**

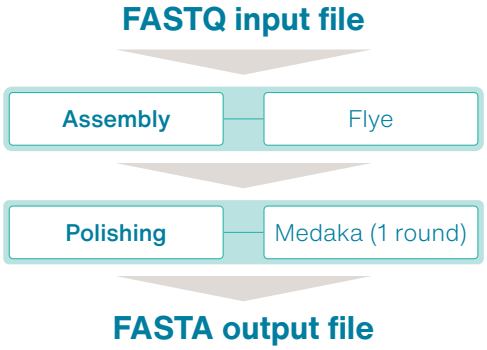
ANALYSIS: selecting an assembly tool

Find out more about data analysis solutions:
nanoporetech.com/analyse

To assemble a human genome, we recommend the third-party *de novo* assembly tool **Flye**[1]. This analysis package represents a complete pipeline, taking raw nanopore reads as input, and producing polished contigs as output. We also advise one round of additional polishing of the assembly with **Medaka**[2].

Assembly can be performed using the on-board PromethION compute — requiring around two days to assemble and polish a single human genome.

These analysis tools can be found on GitHub.



Find out more at: **nanoporetech.com/assembly**

Oxford NANOPORE Technologies

Twitter: @nanopore
www.nanoporetech.com

References:
1. Kolmogorov, M. et al. Nat. Biotechnol. 37:540-546 (2019).
2. Oxford Nanopore Technologies. Medaka. Software available at: https://github.com/nanoporetech/medaka

WF_1066(EN)_V3_14Dec2022